Cluster mass profiles from weak lensing:

shear vs. magnification information

Peter Schneider, Lindsay King & Thomas Erben

Max-Planck-Institut für Astrophysik
Postfach 1523, D-85740 Garching, Germany
e-mail: peter, lindsay, erben@mpa-garching.mpg.de

Abstract

A massive foreground cluster lens changes the shapes (shear effect) and number density (magnification effect) of the faint background galaxy population. In this paper we investigate how the shear, magnification and combined information can be used to constrain cluster mass profiles in the weak lensing regime. We develop maximum likelihood techniques to make quantitative predictions for each of the methods. Our analytic results are checked against Monte Carlo simulations.

In general, we find that the shear method is superior to the magnification method. However, the magnification information complements the shear information if the former has accurate external calibration. For the magnification method, we discuss the effects of random and systematic uncertainties in the background galaxy counts.

1 Introduction

Weak gravitational lensing has been recognized as a powerful tool to investigate the mass and mass distribution of clusters of galaxies (e.g., Webster 1985; Tyson et al. 1990; Kochanek 1990). In their pioneering paper, Kaiser & Squires (1993) pointed out that the tidal gravitational field of the cluster, as measured from the distortion of image shapes of the faint background galaxy population, can be used to reconstruct the two-dimensional projected mass profiles of clusters, without referring to a parametrized mass model. This method, modified in various ways later (e.g., Kaiser 1995; Schneider 1995; Seitz & Schneider 1995; Bartelmann et al. 1996; Seitz & Schneider 1996; Squires & Kaiser 1996; Seitz & Schneider 1997, 1998; Seitz, Schneider & Bartelmann 1998; Lombardi & Bertin 1998a, 1998b) has been applied to more than a dozen clusters up to now (e.g., Fahlman et al. 1994; Smail et al. 1995; Squires et al. 1996a, b; Seitz et al. 1996; Luppino & Kaiser 1997; Clowe et al. 1998; Hoekstra et al. 1998; Kaiser et al. 1999). One of the main results from these studies is that in many clusters the projected mass distribution closely follows the distribution of luminous cluster galaxies. For some of these clusters, very large mass-to-light ratios have been reported, e.g., for MS1224.7+2007, this ratio has been determined by two independent studies to be larger than $\sim 800h$ in solar units (Fahlman et al. 1994; Fischer 1999).

One of the main difficulties of these studies is the existence of the so-called mass-sheet degeneracy (Falco et al. 1985; Schneider & Seitz 1995) which states that the image shapes are unchanged if the surface mass distribution $\kappa(\boldsymbol{\theta})$ (defined in the usual way) is replaced by $\kappa(\boldsymbol{\theta}) \to \lambda\kappa(\boldsymbol{\theta}) + (1-\lambda)$. In particular, this transformation keeps the critical curves of the lens mapping fixed, and thus the location of the giant arcs and arclets. The choice of the constant λ in actual mass reconstructions is largely arbitrary. Provided the observed data field is sufficiently large, one might argue that the surface mass density has decreased to a value close to zero far away from the cluster center, thereby fixing λ. However, for those investigations where the available data field is quite small, the invariance transformation implies a substantial uncertainty in the mass estimates.

The mass sheet degeneracy affects the magnification as $\mu(\boldsymbol{\theta}) \to \mu(\boldsymbol{\theta})/\lambda^2$; hence, if magnification information can be obtained, the degeneracy can be broken. Two approaches to measuring the magnification have been presented in the literature: Broadhurst et al. (1995) considered the effect of magnification on the local number counts of faint background galaxies, employing the so-called magnification bias effect. Since the number counts of very faint galaxies are flatter than the critical slope (defined below), magnification leads to a depletion of the local number counts near the cluster center. This effect has been observed in at least two clusters (Fort et al. 1997; Taylor et al. 1998). Alternatively, Bartelmann & Narayan (1995) have suggested using the size of faint galaxies at fixed surface brightness as a measure for the magnification, making use of the fact that gravitational light deflection conserves the surface brightness. In addition, further measures of the weak lensing magnification have been proposed, such as the change in the redshift distribution of galaxies in a flux-limited sample (Broadhurst et al. 1995), or the effect on the two-point angular correlation function of distant galaxies (Moessner et al. 1998).

Whereas the assumption underlying the cluster mass reconstruction from image ellipticities (the shear method, in what follows), namely that the intrinsic orientation of source galaxies are randomly distributed – or that there is no preferred direction of alignment, is simple and predicts that the mean ellipticity of galaxy images in a small region of the cluster equals the reduced shear, the magnification techniques need an external calibration: for the first of the methods mentioned above, the number density of background galaxies in a 'blank field' as a function of limiting magnitude (and other selection criteria, such as surface brightness etc.) needs to be known, and for the second, the mean size of galaxies as a function of surface brightness is required. Uncertainties in this calibration information immediately translates into uncertainties in the derived magnification. In addition, naive estimates of the accuracy of shear and magnification measurements (see Sect. 2) show that the shear method appears to be superior.

Nevertheless, Broadhurst (1999) claimed that the magnification information can, at least in some cases, yield a more accurate determination of the projected radial mass profiles of clusters. In this paper, we investigate the substance of this claim in more detail. In Sect. 2 we briefly summarize the shear and magnification methods, and provide a rough estimate of their respective accuracy in the determination of local lens parameters. Sect. 3 describes a maximum-likelihood method for the use of shear and magnification information to determine mass profiles. The ensemble-averages of the likelihood functions are derived in Sect. 4; later, these are used to determine the characteristic errors of parameter estimates. In Sect. 5 we describe the numerical simulations used to substantiate our analytic results, and the lens models that are considered. We present the results from the likelihood analysis in Sect. 6; when the unlensed background number density is known, we find that the shapes of the likelihood contours are different in both methods. In particular, for a wide range of situations, the slope of the mass profile is ill-determined from the shear method, and better constrained by the magnification method. In Sect. 6 we also consider the influence of uncertainty in the number density on the likelihood analysis. We summarize our results and conclude in Sect. 7.

2 Shear and magnification methods

We use standard lensing notation; i.e., $\kappa(\boldsymbol{\theta})$ denotes the dimensionless surface mass density of the deflector – which is well defined if we assume that all source galaxies are at the same distance, an assumption which is reasonable for low-redshift clusters ($z_{\rm d} \lesssim 0.25$), and for many purposes not a bad approximation even for medium-redshift lenses. The deflection potential $\psi(\boldsymbol{\theta})$ is related to κ through a Poisson-like equation, $\nabla^2\psi = 2\kappa$. The trace-less part of the Hessian of ψ describes the tidal gravitational field, which we summarize in the complex shear $\gamma = \gamma_1 + {\rm i}\gamma_2 = (\psi_{,11} - \psi_{,22})/2 + {\rm i}\psi_{,12}$, where indices separated by a comma denote partial derivatives with respect to the position $\boldsymbol{\theta}$ on the sky. The magnification of an image is the inverse of the Jacobian determinant of the lens equation, $\mu(\boldsymbol{\theta}) = [\det\mathcal{A}(\boldsymbol{\theta})]^{-1}$, with $\det\mathcal{A} = (1-\kappa)^2 - |\gamma|^2$.

The shear method is based on the transformation between source and image ellipticity due to the tidal field of the deflector. Let $\epsilon^{\rm s}$ denote the intrinsic ellipticity of the source, and ϵ the ellipticity of the observed image. Here, we use the ellipticity parameter as defined in, e.g., Seitz & Schneider (1997), such that ϵ is a complex number whose modulus, in the case of elliptical isophotes with axis ratio r is $|\epsilon| = (1-r)/(1+r)$, and whose phase is twice the position angle of the major axis. The locally linearized lens equation yields the transformation between source and image ellipticity,

$$\epsilon = \frac{\epsilon^{\rm s} + g}{1 + g^*\epsilon^{\rm s}} \; , \tag{1}$$

where $g = \gamma/(1-\kappa)$ is the complex reduced shear, and the asterisk denotes complex conjugation. The transformation (1) is valid only in the non-critical parts of the lens, i.e., where $\det\mathcal{A} > 0$, to which we will restrict the discussion here. Let $p_\epsilon^{\rm s}(\epsilon^{\rm s})\,{\rm d}^2\epsilon^{\rm s}$ be the probability that the source ellipticity lies within ${\rm d}^2\epsilon^{\rm s}$ of $\epsilon^{\rm s}$, then the expectation value of the image ellipticity is

$$\langle\epsilon\rangle := \int {\rm d}^2\epsilon^{\rm s}\; p_\epsilon^{\rm s}(\epsilon^{\rm s})\,\epsilon(\epsilon^{\rm s}) = g \; , \tag{2}$$

as shown by Schramm & Kayser (1995) and Seitz & Schneider (1997). This result implies that the ellipticity of each galaxy image provides an unbiased estimate of the local reduced shear (it should be noted here that the reduced shear is unchanged under the mass-sheet invariance transformation).

The magnification method based on the local number counts of background galaxies arises from the magnification bias (e.g., Canizares 1982) which states that the local cumulative number counts $n(\theta; S)$ above a flux limit S are related to the unlensed counts $n_0(S)$ by

$$n(\theta; S) = \frac{1}{\mu} n_0 \left(\frac{S}{\mu}\right) \; . \qquad (3)$$

If we assume that the number counts follow (locally) a power law of the form $n_0 \propto S^{-\beta}$, then

$$n(\theta) = n_0 \, \mu^{\beta-1} \qquad (4)$$

at any fixed flux threshold. This implies that if the intrinsic counts are flatter than 1, then the lensed counts will be reduced relative to the unlensed ones. This number depletion is the signature of lensing in the magnification method considered here.

Let us briefly consider the signal-to-noise of a shear and a magnification detection by those respective methods. The noise in the former case is due to the intrinsic ellipticity distribution, whereas it is at least Poisson noise for the magnification method. If the source galaxies have a significant angular correlation, the resulting noise will be larger than the naive Poisson estimate used below; we consider this in Sect. 6.3.

Thus, assume that in some (small) solid angle the shear is constant, and that there are N_γ background galaxy images for which the ellipticity can be measured. The signal is g, and the noise is $\sigma_\epsilon/\sqrt{N_\gamma}$, where σ_ϵ is the ellipticity dispersion. Thus, the signal-to-noise for a shear detection is

$$\left(\frac{\mathrm{S}}{\mathrm{N}}\right)_\gamma = \frac{|g|\sqrt{N_\gamma}}{\sigma_\epsilon} \; . \qquad (5)$$

Similarly, assume that the magnification is constant over a solid angle and that, in the absence of lensing, N_μ galaxies would be expected in that region. Then, the magnification signal is $|\Delta N| = |\mu^{\beta-1} - 1| N_\mu$, and the noise is $\sqrt{N_\mu}$, ignoring the change of the number density in the noise estimate. Thus,

$$\left(\frac{\mathrm{S}}{\mathrm{N}}\right)_\mu = |\mu^{\beta-1} - 1|\sqrt{N_\mu} \; . \qquad (6)$$

Specializing now to the case of weak lensing, i.e., $\kappa \ll 1$, $|\gamma| \ll 1$, so that $\gamma \approx g$, the magnification becomes $\mu \approx 1+2\kappa$, and a first-order expansion yields $|\Delta N| \approx 2\kappa|1-\beta|N_\mu$. Then, the ratio of the signal-to-noise estimates becomes

$$\frac{(\mathrm{S/N})_\gamma}{(\mathrm{S/N})_\mu} = \frac{|\gamma|}{\kappa} \, \frac{1}{2\sigma_\epsilon|1-\beta|} \sqrt{\frac{N_\gamma}{N_\mu}} \; . \qquad (7)$$

In a typical situation, the magnitudes of the shear and the surface mass density are similar (they are exactly the same for a singular isothermal sphere), and so the first factor in (7) is of order unity. The slope of the galaxy number counts is flattest for the faintest flux limits and for redder colors, and for these, $\beta \gtrsim 0.5$. The variance of the intrinsic ellipticity is of order $\sigma_\epsilon \approx 0.2$, and so the second factor in (7) is of order 5. The last factor can be expected to be smaller than unity, since the flux limit for counts can be fainter than that for which ellipticity measurements are still possible, but in order to make up for the second factor, the number density used for the counts would need to be 25 times higher than for the shear estimates, which at the assumed slope of the number counts would correspond to about 6 magnitudes! This of course is unrealistic – the ratio of N_μ over N_γ is more typically of order 3 or 4 – and we therefore conclude from this simple analysis that the shear method is considerably more sensitive than the magnification method based on number counts. The other magnification method based on images sizes yields a signal-to-noise ratio comparable to the shear method, but since this effect has not been observed up to now, we shall not consider this method in this paper.

3 Likelihood functions for shear and magnification

We shall now consider the following situation: in a solid angle (e.g., around a cluster), one observes N_μ galaxy images at positions $\boldsymbol{\theta}_i$, $1 \le i \le N_\mu$, down to a given flux threshold. In addition, one observes N_γ galaxy images at positions $\boldsymbol{\vartheta}_i$, $1 \le i \le N_\gamma$, for which an ellipticity can be measured; the two sets of galaxy images can have objects in common. If we now assume that the mass distribution of the lens is described by a parametrized model with parameters π_i, $1 \le i \le M$, what are the best-fitting parameters for the observation?

We solve this problem by defining a likelihood function and then maximizing it. The likelihood is defined as the probability of obtaining the observables given the model parameters π_i, and is considered as the probability for the model parameters π_i given the observables (see, e.g., Press et al. 1992 for a discussion on this point).

Starting with the magnification effect, the probability of observing N_μ galaxy images at positions $\boldsymbol{\theta}_i$ can be factorized in the probability of observing a total of N_μ galaxies, and in the probability that, given that there are N_μ galaxies, they are located at $\boldsymbol{\theta}_i$. For a lens model with parameters π_i, the expected number of galaxy images is

$$\langle N_\mu \rangle = n_\mu \int \mathrm{d}^2\theta \, [\mu(\boldsymbol{\theta})]^{\beta-1} \; , \tag{8}$$

as follows from (4); we have used n_μ to denote the unlensed number density, to distinguish it from the number density of galaxies n_γ for which an ellipticity can be measured, and the integral extends over the solid angle of the observations. In practical applications, this area can be quite complicated since accurate photometry of very faint objects is impossible near brighter objects – like cluster galaxies – which have to be masked out (see Taylor et al. 1998 for an impressive illustration of this effect). The probability of finding N_μ galaxies in the data field follows a Poisson distribution (if, as we shall assume, the source galaxies are uncorrelated), and the probability that the i-th galaxy is located at $\boldsymbol{\theta}_i$ is proportional to $n(\boldsymbol{\theta}_i) \propto [\mu(\boldsymbol{\theta}_i)]^{\beta-1}$. Then,

$$\mathcal{L}_\mu = P(N_\mu; \langle N_\mu \rangle) \prod_{i=1}^{N_\mu} \frac{[\mu(\boldsymbol{\theta}_i)]^{\beta-1}}{\int \mathrm{d}^2\theta \, [\mu(\boldsymbol{\theta})]^{\beta-1}} \; , \tag{9}$$

with

$$P(N; \langle N \rangle) = \frac{\langle N \rangle^N}{N!} \exp(-\langle N \rangle) \; . \tag{10}$$

One can now see that the denominator in the product on eq. (9) is $\langle N_\mu \rangle / n_\mu$, and that therefore the factor $\langle N_\mu \rangle^{N_\mu}$ of the Poisson distribution is cancelled. In addition, the factors $N_\mu!$ and $n_\mu^{N_\mu}$ do not depend on the lens model and yield just an irrelevant multiplicative constant in the likelihood; therefore we shall drop this factor. Then, the log-likelihood function becomes

$$\ell_\mu := -\ln \mathcal{L}_\mu = n_\mu \int \mathrm{d}^2\theta\, [\mu(\boldsymbol{\theta})]^{\beta-1} + (1-\beta) \sum_{i=1}^{N_\mu} \ln \mu(\boldsymbol{\theta}_i) \; . \tag{11}$$

The best fitting model is now obtained by minimizing ℓ_μ with respect to the model parameters which are contained in the function $\mu(\boldsymbol{\theta})$.

Next we turn to the likelihood function for the shear estimate. For that, we need the probability that the observed ellipticity at position $\boldsymbol{\vartheta}_i$ is ϵ, given that the reduced shear is $g(\boldsymbol{\vartheta}_i)$. Assuming a Gaussian intrinsic ellipticity distribution,

$$p_\epsilon^{\mathrm{s}}(\epsilon^{\mathrm{s}}) = \frac{\exp\left(-|\epsilon^{\mathrm{s}}|^2/\sigma_\epsilon^2\right)}{\pi\sigma_\epsilon^2\left(1-\exp\left(-1/\sigma_\epsilon^2\right)\right)} \tag{12}$$

the observed probability distribution can be obtained from

$$p_\epsilon(\epsilon|g) = p_\epsilon^{\mathrm{s}}(\epsilon^{\mathrm{s}}(\epsilon|g)) \left| \frac{\partial^2\epsilon^{\mathrm{s}}}{\partial^2\epsilon} \right| = p_\epsilon^{\mathrm{s}}(\epsilon^{\mathrm{s}}(\epsilon|g)) \frac{(|g|^2-1)^2}{|\epsilon g^* - 1|^4} \; ,$$

as in Geiger & Schneider (1998). The corresponding probability density $p_\epsilon(\epsilon_i)$ for each lensed galaxy in the catalog is determined, and the likelihood and log-likelihood functions are given by

$$\mathcal{L}_\gamma = \prod_{i=1}^{N_\gamma} p_\epsilon(\epsilon_i|g(\boldsymbol{\vartheta}_i)) \; ; \qquad \ell_\gamma = -\sum_{i=1}^{N_\gamma} \ln p_\epsilon(\epsilon_i|g(\boldsymbol{\vartheta}_i)).$$

We shall follow a somewhat simpler road in the analytic treatment by assuming that the observed ellipticity distribution is a Gaussian; this is a reasonable assumption if the intrinsic ellipticity distribution is a Gaussian and if the reduced shear is not too close to unity. The expectation value of ϵ is g, and the dispersion in the two directions parallel and perpendicular to the shear are the same (see Appendix of Geiger & Schneider 1999) and given by

$$\sigma^2 = 2\sigma_\perp^2 = 2\sigma_\parallel^2 = 2\pi(1-|g|^2)^2 \int_0^1 \mathrm{d}|\epsilon^{\mathrm{s}}| \; \frac{|\epsilon^{\mathrm{s}}|^3\, p_\epsilon^{\mathrm{s}}(|\epsilon^{\mathrm{s}}|)}{1-|\epsilon^{\mathrm{s}}|^2 g^2} \; , \tag{13}$$

as will be shown in the Appendix. Provided $|g|$ is much smaller than unity, or that the intrinsic probability distribution does not extend significantly out to $|\epsilon^{\mathrm{s}}| \sim 1$, the denominator in the integral can be replaced by unity, for which the dispersion of the observed ellipticities becomes

$$\sigma \approx (1-|g|^2)\, \sigma_\epsilon \; . \tag{14}$$

In the analytical treatment of Sect. 4, we shall use the approximation (13) for simplicity (for $\sigma_\epsilon = 0.2$, the approximation for σ differs from the true value by less than 5% for $|g| \leq 1$, and less than 1% for $|g| \leq 0.4$). The likelihood function for the shear method is then simply

$$\mathcal{L}_\gamma = \prod_{i=1}^{N_\gamma} \frac{1}{\pi\sigma^2[g(\boldsymbol{\vartheta}_i)]} \exp\left(-\frac{|\epsilon_i - g(\boldsymbol{\vartheta}_i)|^2}{\sigma^2[g(\boldsymbol{\vartheta}_i)]}\right) \; ,$$

and the corresponding log-likelihood function becomes, dropping irrelevant additive constants,

$$\ell_\gamma = \sum_{i=1}^{N_\gamma} \left[\frac{|\epsilon_i - g(\boldsymbol{\vartheta}_i)|^2}{\sigma^2[g(\boldsymbol{\vartheta}_i)]} + 2\ln\sigma[g(\boldsymbol{\vartheta}_i)] \right] \; . \tag{15}$$

Again, the best-fitting lens model is obtained by minimizing ℓ_γ with respect to the model parameters, which enter g and thus also σ.

The combined shear and magnification likelihood is then obtained by just multiplying the respective likelihoods, or adding the log-likelihoods,

$$\ell_{\rm tot} = \ell_\mu + \ell_\gamma \; . \tag{16}$$

4 Ensemble-averaged log-likelihood

For a given set of observables, the likelihood functions defined in the previous section can be calculated; the minimum yields the best-fitting model parameters, and the width of the likelihood function yields the confidence region for the parameter estimate from this data set. In order to determine the characteristic errors of parameter estimates, we consider here the ensemble-averaged log-likelihood function.

Given a lens model – hereafter called the true model, and characterized by subscripts 't' – the ensemble-average of a quantity X is provided by

$$\langle X \rangle := \sum_{N=0}^{\infty} P(N; \langle N \rangle_{\rm t}) \left[\prod_{i=1}^{N} \int {\rm d}^2\theta_i \; p_{\rm t}(\boldsymbol{\theta}_i) \int {\rm d}^2\epsilon_i \; p_{\rm t}(\epsilon_i) \right] X \; , \tag{17}$$

where the first term averages over the probability to find N galaxies when the expectation value from the true model is $\langle N \rangle_{\rm t}$, the second term is the integration over the probability that galaxy number i lies at $\boldsymbol{\theta}_i$, and the final term integrates over the ellipticity distribution. To see how this procedure works, we start with the log-likelihood function for the magnification method, i.e., setting $X = \ell_\mu$. Since ℓ_μ does not depend on the ellipticity of the images, the final integration in (17) yields simply a factor one. The probability for the positions of the galaxies is

$$p_{\rm t}(\boldsymbol{\theta}_i) = \frac{[\mu_{\rm t}(\boldsymbol{\theta}_i)]^{\beta-1}}{\int {\rm d}^2\theta \, [\mu_{\rm t}(\boldsymbol{\theta})]^{\beta-1}} \; . \tag{18}$$

Then,

$$\langle \ell_\mu \rangle = \sum_{N_\mu=0}^{\infty} P(N_\mu; \langle N_\mu \rangle_{\rm t}) \left[\prod_{i=1}^{N_\mu} \int {\rm d}^2\theta_i \; p_{\rm t}(\boldsymbol{\theta}_i) \right] \ell_\mu \; , \tag{19}$$

with

$$\langle N_\mu \rangle_{\rm t} = n_\mu \int {\rm d}^2\theta \; [\mu_t(\boldsymbol{\theta})]^{\beta-1} \; . \tag{20}$$

The first term in the log-likelihood function (11) depends neither on $\boldsymbol{\theta}_i$ nor on the actual number N_μ of galaxies, and thus the averaging operator in (19) leaves this term unchanged. The second term of ℓ_μ consists of a sum over individual galaxy images; therefore, for the j-th galaxy image, only the term $i = j$ in the spatial averaging in

(19) contributes, the other integrations just yield a factor one. Since there are N_μ equal terms, one obtains

$$\langle \ell_\mu \rangle = n_\mu \int \mathrm{d}^2\theta \, [\mu(\boldsymbol{\theta})]^{\beta-1} + (1-\beta) \sum_{N_\mu=0}^{\infty} P(N_\mu; \langle N_\mu \rangle_\mathrm{t}) \frac{N_\mu}{\int \mathrm{d}^2\theta \, [\mu_\mathrm{t}(\boldsymbol{\theta})]^{\beta-1}} \int \mathrm{d}^2\theta \, [\mu_\mathrm{t}(\boldsymbol{\theta})]^{\beta-1} \ln \mu(\boldsymbol{\theta}) \, . \tag{21}$$

The sum over N_μ can now be performed, and using (20) we obtain

$$\langle \ell_\mu \rangle = n_\mu \int \mathrm{d}^2\theta \, [\mu(\boldsymbol{\theta})]^{\beta-1} + n_\mu (1-\beta) \int \mathrm{d}^2\theta \, [\mu_\mathrm{t}(\boldsymbol{\theta})]^{\beta-1} \ln \mu(\boldsymbol{\theta}) \, . \tag{22}$$

From the derivative of $\langle \ell_\mu \rangle$ with respect to the parameter π_i,

$$\frac{\partial \langle \ell_\mu \rangle}{\partial \pi_i} = (1-\beta) n_\mu \int \mathrm{d}^2\theta \, \frac{1}{\mu(\boldsymbol{\theta})} \frac{\partial \mu(\boldsymbol{\theta})}{\partial \pi_i} \left\{ [\mu_\mathrm{t}(\boldsymbol{\theta})]^{\beta-1} - [\mu(\boldsymbol{\theta})]^{\beta-1} \right\} \, , \tag{23}$$

one immediately sees that $\langle \ell_\mu \rangle$ attains a minimum when the π_i attain their true values, i.e., when $\mu = \mu_\mathrm{t}$.

If one assumes that the likelihood function behaves approximately like a Gaussian near its maximum, then the second partial derivatives of $\langle \ell_\mu \rangle$ contain the information about the confidence region, and so we define

$$V_{ij}^\mu := \left. \frac{\partial^2 \langle \ell_\mu \rangle}{\partial \pi_i \partial \pi_j} \right|_{\pi=\pi_\mathrm{t}} = (1-\beta)^2 n_\mu \int \mathrm{d}^2\theta \, [\mu_\mathrm{t}(\boldsymbol{\theta})]^{\beta-3} \left(\frac{\partial \mu(\boldsymbol{\theta})}{\partial \pi_i} \frac{\partial \mu(\boldsymbol{\theta})}{\partial \pi_j} \right)_{|\pi=\pi_\mathrm{t}} \, , \tag{24}$$

where the values of the π derivatives are evaluated at the true parameter values.

Next, we consider the ensemble average of the log-likelihood function for the shear. For that we have to specify the probability distribution for the image ellipticities for the true model. As mentioned before, we shall use a Gaussian with mean $g_\mathrm{t}(\boldsymbol{\vartheta})$, the true reduced shear at position $\boldsymbol{\vartheta}$, and dispersion given by (13), again with $g = g_\mathrm{t}$, and denoted by σ_t. Then, from (15) and (17), we see that, since ℓ_γ is a sum over galaxy images, the integration operators work term by term, and so we get N_γ identical terms from the integrations. This then immediately allows one to perform the sum over N_γ, as before, so that

$$\langle \ell_\gamma \rangle = \langle N_\gamma \rangle_\mathrm{t} \int \mathrm{d}^2\vartheta \, \frac{[\mu_\mathrm{t}(\boldsymbol{\vartheta})]^{\beta-1}}{\int \mathrm{d}^2\theta \, [\mu_\mathrm{t}(\boldsymbol{\theta})]^{\beta-1}} \times \int \mathrm{d}^2\epsilon \frac{1}{\pi \sigma_\mathrm{t}^2(\boldsymbol{\vartheta})} \exp\left(-\frac{|\epsilon - g_\mathrm{t}(\boldsymbol{\vartheta})|^2}{\sigma_\mathrm{t}^2(\boldsymbol{\vartheta})} \right) \left(\frac{|\epsilon - g(\boldsymbol{\vartheta})|^2}{\sigma^2(\boldsymbol{\vartheta})} + 2 \ln \sigma(\boldsymbol{\vartheta}) \right) \, , \tag{25}$$

and $\langle N_\gamma \rangle_\mathrm{t}$ is defined in analogy to (20), with n_μ replaced by n_γ. The ϵ-integration is readily performed, and one obtains

$$\langle \ell_\gamma \rangle = n_\gamma \int \mathrm{d}^2\vartheta \, [\mu_\mathrm{t}(\boldsymbol{\vartheta})]^{\beta-1} \left(\frac{|g(\boldsymbol{\vartheta}) - g_\mathrm{t}(\boldsymbol{\vartheta})|^2 + \sigma_\mathrm{t}^2(\boldsymbol{\vartheta})}{\sigma^2(\boldsymbol{\vartheta})} + 2 \ln \sigma(\boldsymbol{\vartheta}) \right) \, . \tag{26}$$

By taking a first partial derivative of $\langle \ell_\gamma \rangle$ with respect to the parameter π_i,

$$\frac{\partial \langle \ell_\gamma \rangle}{\partial \pi_i} = n_\gamma \int \mathrm{d}^2\vartheta \; [\mu_\mathrm{t}(\boldsymbol{\vartheta})]^{\beta-1} \, \frac{\partial Y}{\partial \pi_i} \; , \tag{27}$$

with

$$\frac{\partial Y}{\partial \pi_i} = \frac{2}{\sigma^2} \mathcal{R}\mathrm{e}\left[\frac{\partial g}{\partial \pi_i}(g^* - g_\mathrm{t}^*)\right] - \frac{2\left[|g-g_\mathrm{t}|^2 + (\sigma_\mathrm{t}^2 - \sigma^2)\right]}{\sigma^3} \frac{\partial \sigma}{\partial \pi_i} \; , \tag{28}$$

we see that $\langle \ell_\gamma \rangle$ takes a minimum if $\pi = \pi_\mathrm{t}$, so that $g = g_\mathrm{t}$ and therefore $\sigma = \sigma_\mathrm{t}$.

From the ensemble-averaged log-likelihood function, we estimate the dispersion of the parameters one obtains from a single realization of the observables by using the fact that asymptotically (in the limit of infinitely many data points) the distribution of $2\Delta\ell$ tends to a χ^2_M distribution with M being the number of model parameters. Since the number of galaxies is finite, the likelihood function is not Gaussian, and this estimate can only be approximate. We shall demonstrate with simulations that it provides a very good approximation indeed.

5 Simulations and models

The situation we shall consider is that of an axially-symmetric mass distribution for the cluster. We assume that the central part of the cluster is a strong lens where arcs and multiple images are found, and from their modelling the Einstein radius θ_E of the cluster is known. We then assume that the weak lensing analysis is performed in an annulus with inner and outer radii θ_in and θ_out, with $\theta_\mathrm{in} > \theta_\mathrm{E}$. If $\beta < 1$ (we shall assume throughout the paper that $\beta = 0.5$), the number density of galaxies in the annulus is always smaller than the unlensed number density n_0, since $\mu \geq 1$ outside the critical curve (magnification theorem, see Schneider 1984). Hence, we can use the rejection method, as explained below.

5.1 Description of the simulations

Given a specific lens model, galaxy images are distributed in the following way: First, the expected number of galaxies in the annulus in the absence of lensing is calculated, $\langle N_0 \rangle = n_0 \pi (\theta_\mathrm{out}^2 - \theta_\mathrm{in}^2)$, and a number N_0 is drawn from a Poisson distribution with mean $\langle N_0 \rangle$. Then, these N_0 galaxies are provisionally distributed in angular radius, by drawing a random number ξ uniformly distributed in $[0,1]$, and assigning a radius of $\theta = \sqrt{\xi(\theta_\mathrm{out}^2 - \theta_\mathrm{in}^2) + \theta_\mathrm{in}^2}$. Then, a second uniform deviate $\eta \in [0,1]$ is drawn, and a galaxy at θ is put into the final 'catalog' only if $[\mu(\theta)]^{\beta-1} \geq \eta$; otherwise it is discarded. With this rejection method, the distribution of galaxies follows the distribution (18). For each galaxy in the 'catalog', a source ellipticity ϵ^s is drawn from a two-dimensional Gaussian probability density distribution given by equation (12), and the corresponding image ellipticity is obtained from (1). As a result, we obtain a catalog containing radial positions θ_i and image ellipticities ϵ_i. For such a catalog, and an appropriately parametrized family of lens models, the log-likelihood functions ℓ_μ, ℓ_γ and ℓ_tot can be minimized, resulting in the best-fitting parameters.

The number density of galaxies for the shear and the magnification method can be different; throughout we assume that $n_\gamma = 30\mathrm{arcmin}^{-2}$, and $n_\mu = 120\mathrm{arcmin}^{-2}$ (the former of these values is the typical number density used for shear analysis from deep ground-based data, the latter is similar to those used by Fort et al. 1997 and by

Broadhurst 1999). Furthermore, we shall assume that the Einstein radius of the cluster is $\theta_{\rm E} = 0.5$arcmin, as is approximately the case for the cluster CL0024+17 to which the magnification method has been applied (Fort et al. 1997; Broadhurst 1999). The resulting confidence regions of the parameter estimates depend solely on the ratios $\theta_{\rm in}/\theta_{\rm E}$, $\theta_{\rm out}/\theta_{\rm E}$, and on the products $\theta_E^2 n_\gamma$ and $\theta_E^2 n_\mu$, and so can be easily scaled for different values of the Einstein radius.

5.2 Lens models

We shall consider two families of two-parameter lens models. The first (Family A) is characterized by a power-law profile outside the Einstein radius, i.e.,

$$\kappa(\theta) = a \left(\frac{\theta}{\theta_{\rm E}}\right)^{-q} \quad \text{for } \theta \geq \theta_{\rm E} \; ; \tag{29}$$

hence, a is the surface mass density at the Einstein radius. For example, a singular isothermal sphere would have $a = 0.5$ and $q = 1$. Note that we do not have to specify the mass distribution inside the Einstein radius; it suffices to know that the mean surface mass density inside the Einstein radius is $\bar\kappa = 1$. Using the methods described in Chap. 8 of Schneider et al. (1992; hereafter SEF), we find that

$$\bar\kappa(\theta) = \left(1 - \frac{2a}{2-q}\right)\left(\frac{\theta}{\theta_{\rm E}}\right)^{-2} + \frac{2a}{2-q}\left(\frac{\theta}{\theta_{\rm E}}\right)^{-q} \; , \tag{30}$$

and

$$\gamma(\theta) = \bar\kappa(\theta) - \kappa(\theta) \; . \tag{31}$$

Thus, this family has a 'shape' parameter q and a 'strength' parameter a.

For the second family of lens models (hereafter Family B), we use the model described in Sect. 8.1.5 of SEF, which reads

$$\kappa(\theta) = \kappa_0 \frac{1 + p(\theta/\theta_{\rm c})^2}{[1 + (\theta/\theta_{\rm c})^2]^{2-p}} \; , \tag{32}$$

where $\theta_{\rm c}$ is the core radius, and κ_0 the central surface mass density. For $\theta \ll \theta_{\rm c}$, the surface mass density is nearly constant at κ_0, whereas for $\theta \gg \theta_{\rm c}$ it decreases like a power law, with a slope determined by p. The Einstein radius is given as

$$\theta_{\rm E} = \theta_{\rm c}\sqrt{\kappa_0^{1/(1-p)} - 1} =: \theta_{\rm c}\sqrt{W} \; , \tag{33}$$

and only exists if $\kappa_0 > 1$, a well-known result for centrally-concentrated axially-symmetric lenses. Hence, we can eliminate $\theta_{\rm c}$ in terms of $\theta_{\rm E}$ in (32) and again obtain a two-parameter model (for fixed $\theta_{\rm E}$),

$$\kappa(\theta) = \kappa_0 \frac{1 + pW(\theta/\theta_{\rm E})^2}{[1 + W(\theta/\theta_{\rm E})^2]^{2-p}} \; , \tag{34}$$

which leaves κ_0 and p as the two parameters. Asymptotically, for $\theta \to \infty$, the mass distribution behaves like

$$\kappa \to \kappa_0 \, p \, W^{p-1} (\theta/\theta_{\rm E})^{2p-2} \; .$$

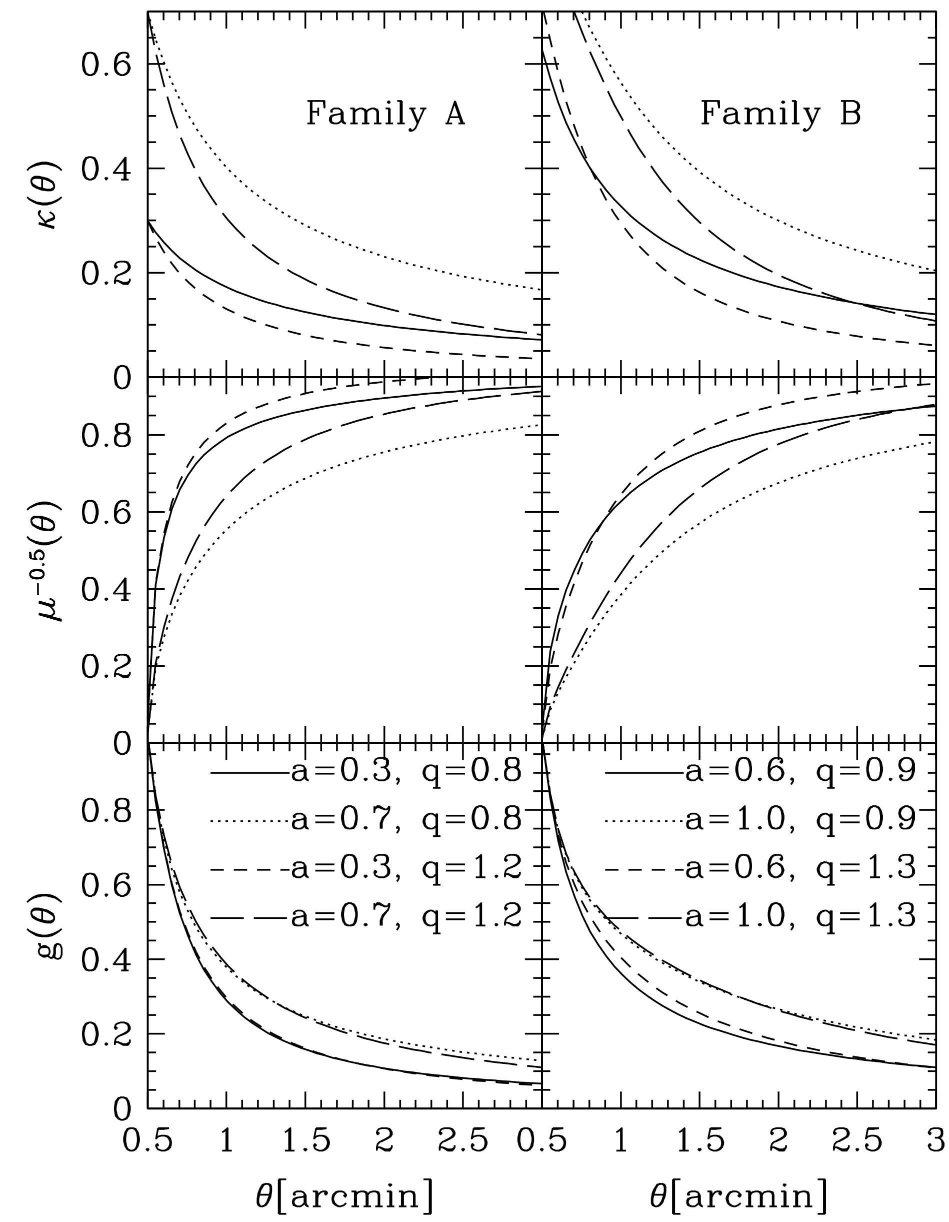

Fig. 1. The radial mass profile κ (upper panels), the magnification signature for the image density depletion, $\mu^{-0.5}$ (middle panels), and the reduced shear g (bottom panels) as a function of angular separation from the cluster center, for four different combinations of the model parameters a and q, as indicated. Left panels are for models of Family A, right panels for Family B

We shall choose the two free parameters in such a way that the asymptotic form of this model is similar to that of Family A. Hence we take $a = \kappa_0 p W^{p-1}$ and $q = 2 - 2p$ as our parameters. Then, $1/W = [2a/(2-q)]^{2/q} - 1$, $\kappa_0 = (1+W)^{q/2}$, and therefore,

$$\kappa(\theta) = \left(\frac{2a}{2-q}\right) \frac{\left(\frac{2a}{2-q}\right)^{2/q} - 1 + \left(\frac{2-q}{2}\right)\left(\frac{\theta}{\theta_{\rm E}}\right)^2}{\left[\left(\frac{2a}{2-q}\right)^{2/q} - 1 + \left(\frac{\theta}{\theta_{\rm E}}\right)^2\right]^{1+q/2}} \; ; \tag{35}$$

accordingly,

$$\bar\kappa(\theta) = \left(\frac{2a}{2-q}\right)\left[\left(\frac{2a}{2-q}\right)^{2/q} - 1 + \left(\frac{\theta}{\theta_{\rm E}}\right)^2\right]^{-q/2} . \tag{36}$$

As before the shear is calculated from (31). The condition $\kappa_0 > 1$, or $W > 0$, translates into

$$a > \frac{2-q}{2} \; . \tag{37}$$

For $a = (2-q)/2$, the resulting lens model coincides with the corresponding one from Family A.

In Fig. 1 we plot, for four different combinations of a and q, the radial mass profile $\kappa(\theta)$, the magnification signature $\mu^{\beta-1}$ for $\beta = 0.5$, and the reduced shear g, for both Families of models. The upper panels show the radial dependence of the surface mass density $\kappa(\theta)$ which are all fairly different over the range of angles plotted. The bottom panels show the radial dependence of the reduced shear; here we see, in particular for Family A, that models with the same amplitude but different slope have nearly the same curves $g(\theta)$; hence, it will be quite challenging to determine the radial slope q for these models from the shear method. The reason for this similarity comes from an accidental cancellation of the slope dependencies of $1-\kappa$ and γ in g for these models. In contrast to the reduced shear, the magnification signal, plotted as $\mu^{\beta-1} = \mu^{-0.5}$ in the middle panels, differs clearly between models with different slopes. Hence, despite the conclusion reached in Sect. 2 on the relative sensitivity of the shear and magnification methods, from the curves in Fig. 1 one might expect that the magnification effect could more easily distinguish between different radial slopes.

6 Results from the likelihood analysis

We now proceed to consider the likelihood functions for the models described in the previous Section, starting with the analytic results in Sect. 6.1. In Sect. 6.2 the results of the numerical simulations are used to substantiate our analytic treatment. We also ask whether it is possible to distinguish between lens Families, using each of the methods. In Sect. 6.3 we consider how the uncertainty in the unlensed background number density influences the magnification likelihood functions.

6.1 Analytic results

In Figs. 2 and 3, contours of constant likelihood are plotted, for $\langle\ell_\gamma\rangle$, $\langle\ell_\mu\rangle$, and $\langle\ell_{\rm tot}\rangle$, for models of Family A and B, respectively. The contours are such that one expects the parameters obtained by minimizing the log-likelihood from realizations of galaxy image observations to lie within the contours in 68.3%, 90%, 95.4%, 99%, 99.73%, and 99.99% percent of all cases. This expectation is based on the approximation mentioned at the

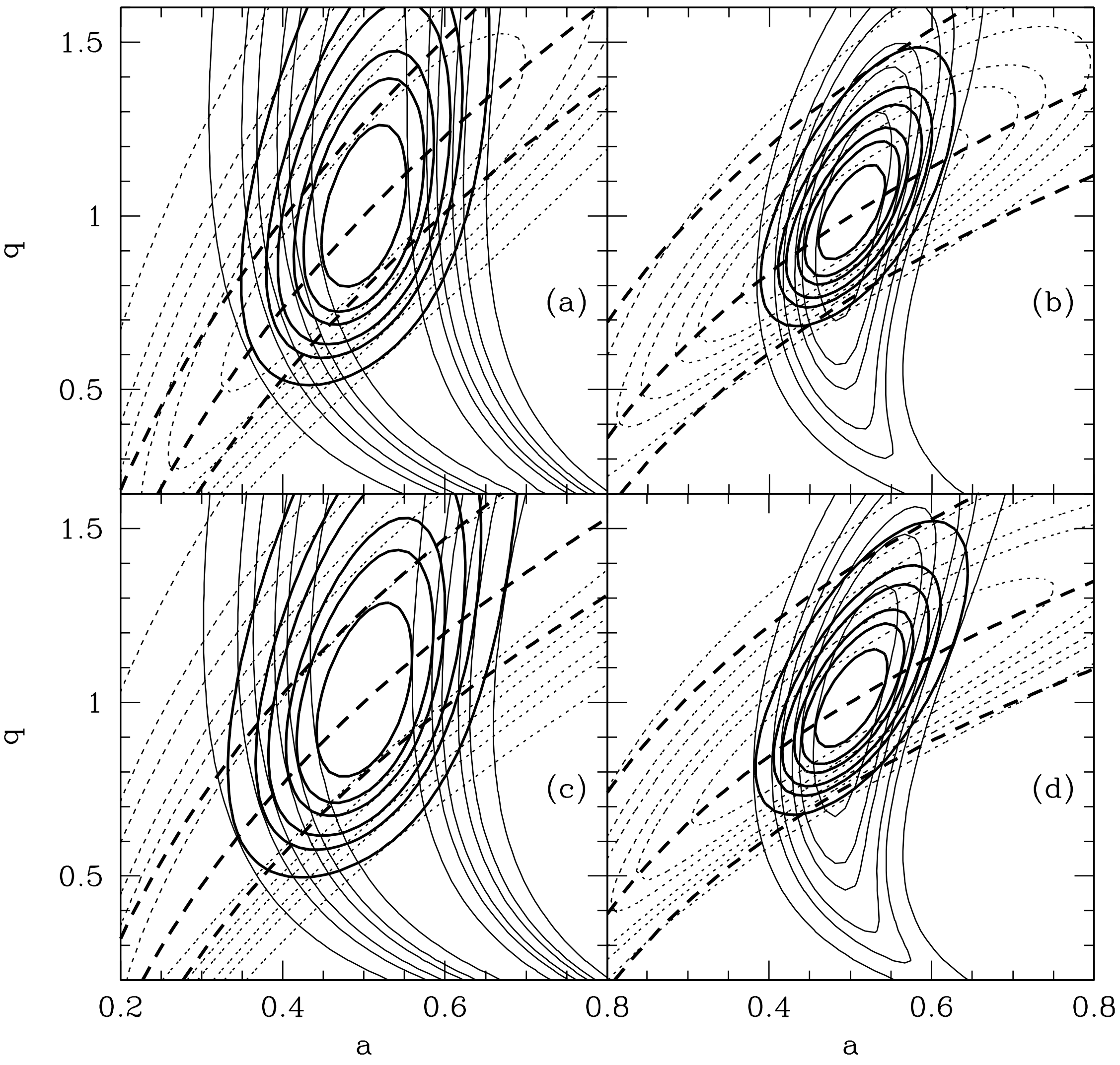

Fig. 2. Contours of constant average likelihood for models of Family A. The thin solid lines are contours of constant $\langle \ell_\gamma \rangle$, the dashed contours correspond to $\langle \ell_\mu \rangle$, and the heavy solid contours to $\langle \ell_{\rm tot} \rangle$. Contours are drawn for $2\Delta\ell = \{2.30, 4.61, 6.17, 9.21, 11.8, 18.4\}$, within which one expects that 68.3%, 90%, 95.4%, 99%, 99.73%, and 99.99% respectively, of parameter estimates from realizations will be enclosed. The input model is described by $a = 0.5$, $q = 1.0$, and corresponds to the minimum of $\langle \ell \rangle$. Parameters are as described in the text, i.e., $n_\mu = 120\,{\rm arcmin}^{-2}$, $n_\gamma = 30\,{\rm arcmin}^{-2}$, $\beta = 0.5$, $\sigma_\epsilon = 0.2$, and the inner and outer radii of the annulus are (a) $\theta_{\rm in} = 0\farcm6$, $\theta_{\rm out} = 2\farcm0$; (b) $\theta_{\rm in} = 0\farcm6$, $\theta_{\rm out} = 4\farcm0$; (c) $\theta_{\rm in} = 0\farcm9$, $\theta_{\rm out} = 2\farcm0$; (d) $\theta_{\rm in} = 0\farcm9$, $\theta_{\rm out} = 4\farcm0$. The three heavy solid curves are curves of constant $\langle N_\mu \rangle$; the middle one is for $\langle N_\mu \rangle$ fixed at the value obtained for the input model, the two others are for changes of $\langle N_\mu \rangle$ by $\pm 5\%$

end of Sect. 4. The various panels in the two figures differ in the choice of the inner and outer radii of the annulus in which the data are assumed to be analyzed.

Considering Fig. 2 first, the first thing to note is that the contours obtained from the shear method are very nearly parallel to the q-axis. This is in accordance with our discussion of Fig. 1 at the end of the previous section where we have seen that the reduced shear depends only weakly on the slope q of the mass profile. Secondly, the direction of elongation of the likelihood contours corresponding to the magnification method is nicely off-set from the q-axis; this implies that the combination of shear and magnifica-

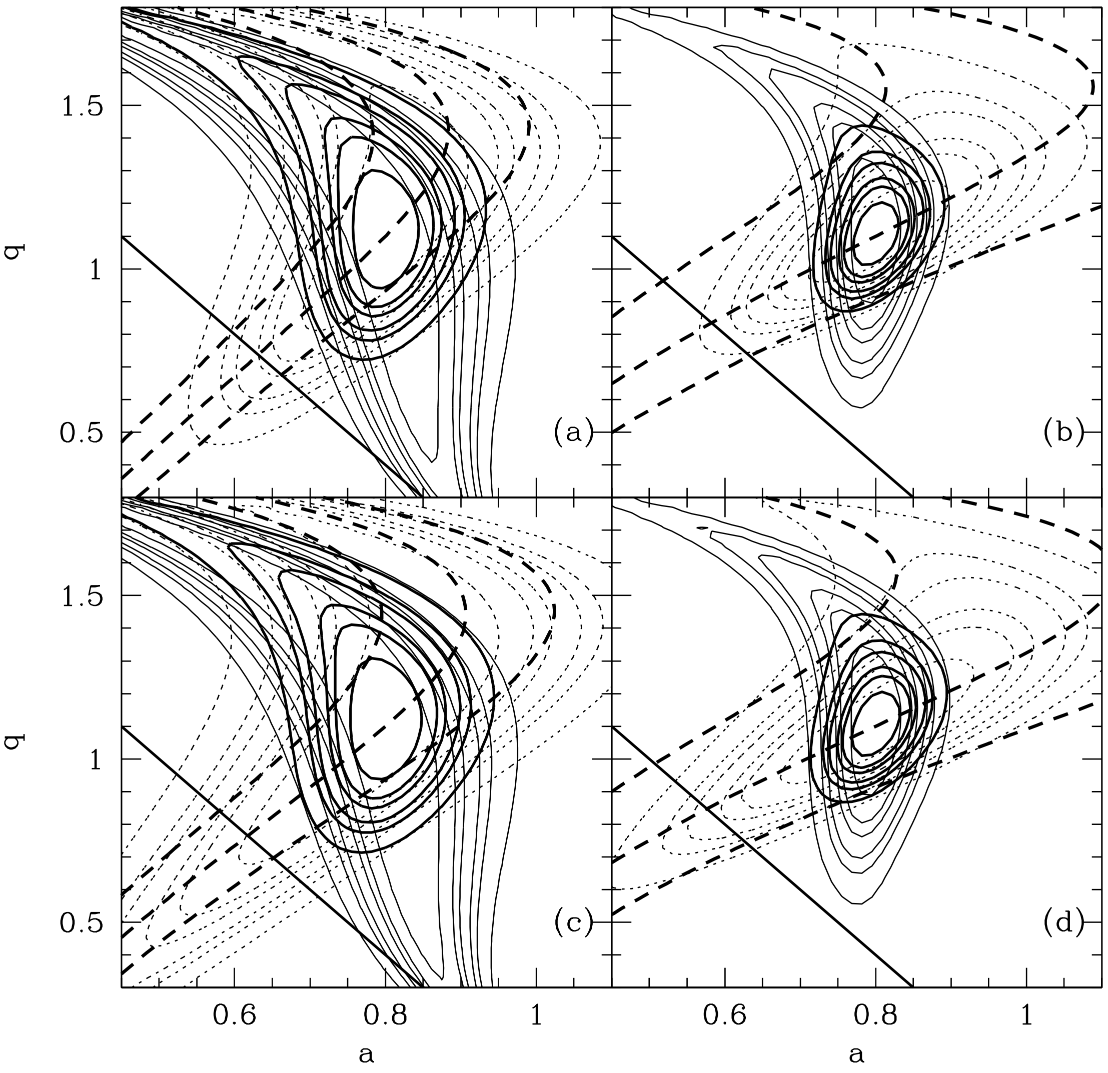

Fig. 3. Same as Fig. 2, but now for models of Family B. The input model is characterized by $a = 0.8$, $q = 1.1$. Inner and outer radii are the same as in Fig. 2. The heavy solid line in each panel denotes the limit in (37); models below this line do not correspond to centrally condensed mass distributions with $\kappa_0 > 1$

tion information is much more powerful than each method individually in distinguishing between the model parameters. This fact is actually demonstrated by the likelihood contours corresponding to $\langle \ell_{\rm tot} \rangle$ which are much more compact than the other two sets of contours. Considering the upper left panel as an example, the 1-σ contour of the magnification method closes within the boundary of the figure, whereas the 1-σ contour of $\langle \ell_\gamma \rangle$ is open both to the top and the bottom; hence, from the magnification alone one can constrain the slope and the amplitude, whereas the shear method yields basically no constraint on the slope. Increasing the outer radius of the annulus, the shear method becomes substantially more sensitive; in the upper right panel of Fig. 2, the areas enclosed by the corresponding contours for the shear and the magnification methods are about the same.

In order to check whether most of the information comes from galaxy images very close to the critical curves at $\theta = \theta_{\rm E} = 0\rlap{.}'5$, we have increased the inner radius of the annulus in the lower two panels. Although the likelihood contours widen slightly, this is not a dramatic effect; in particular, the likelihood contours for the combination of both methods remain basically unchanged. Therefore, most of the signal is not located right at the Einstein radius. This is good news, since there the measurement of galaxy shapes and fluxes may be affected by bright galaxies in the center of the cluster.

For the models of Family B, the situation is qualitatively the same as discussed for Family A. Here, the advantage of increasing the outer radius of the annulus is even larger for the shear method than was the case for the Family A models. This result shows the importance of wide-field images for a quantitative analysis of cluster mass profiles (see Bonnet et al. 1994). As for Family A, little information is lost by increasing the inner radius from $1.2\theta_{\rm E}$ to $1.8\theta_{\rm E}$.

In order to investigate a larger part of parameter space, we consider the behaviour of the log-likelihood function near the minimum. Writing

$$\left\langle \ell \right\rangle (\pi) \approx \left\langle \ell \right\rangle (\pi_{\rm t}) + \frac{1}{2} \frac{\partial^2 \left\langle \ell \right\rangle}{\partial \pi_i \, \partial \pi_j} \, (\Delta\pi)_i \, (\Delta\pi)_j \tag{38}$$

for parameter values close to those of the 'true' parameters, curves of constant $\Delta \left\langle \ell \right\rangle$ are described as ellipses. The '1-σ error ellipse' is described by setting $2\Delta \left\langle \ell \right\rangle = 2.30 \equiv X$. Note that from Figs. 2 and 3 one finds that the confidence contours are not always well approximated by ellipses; nevertheless, the consideration here should provide a qualitative handle on the expected uncertainties in parameter determinations. The matrix of partial derivatives in (38) has been calculated for the magnification method in (24); the corresponding expression for the shear method is too long to be reproduced here. The '1-σ error ellipse' is characterized by three numbers, corresponding to the major and minor axes, and their orientation. In order to translate this into errors Δa and Δq for parameter estimates, we chose the following prescription: We construct the smallest rectangle which encloses the ellipse, and take as errors the half-lengths of its sides. Denoting the matrix in (38) by V, one then obtains

$$\begin{aligned} \frac{\Delta a}{\sqrt{X}} &= \sqrt{\frac{\det V}{V_{11}^2 V_{22}}} + \frac{V_{12}^2}{V_{11}} \sqrt{\frac{1}{V_{22} \det V}} \,, \\ \frac{\Delta q}{\sqrt{X}} &= \sqrt{\frac{\det V}{V_{11} V_{22}^2}} + \frac{V_{12}^2}{V_{22}} \sqrt{\frac{1}{V_{11} \det V}} \,. \end{aligned} \tag{39}$$

This graphical prescription for the error in the two parameters is certainly not optimal, but easy to apply; in cases where the eigendirections of V are close to 'diagonal' (as is the case for the magnification method in Fig. 2), this method can severely overestimate the error.

The errors determined in that way are plotted in Fig. 4, for two choices of the radii of the annulus in which observations are assumed to be available; the small (large) outer radius corresponds to the left (right) panels in Fig. 4. They are plotted as a function of the value of a, for five different values of q, and using the shear (solid curves), magnification (dotted curves) and the combination of both methods (dashed curves). The most obvious point to note is that for the smaller outer radius of the annulus, the magnification method yields smaller errors in q than the shear method, for all values of a and q (see lower left

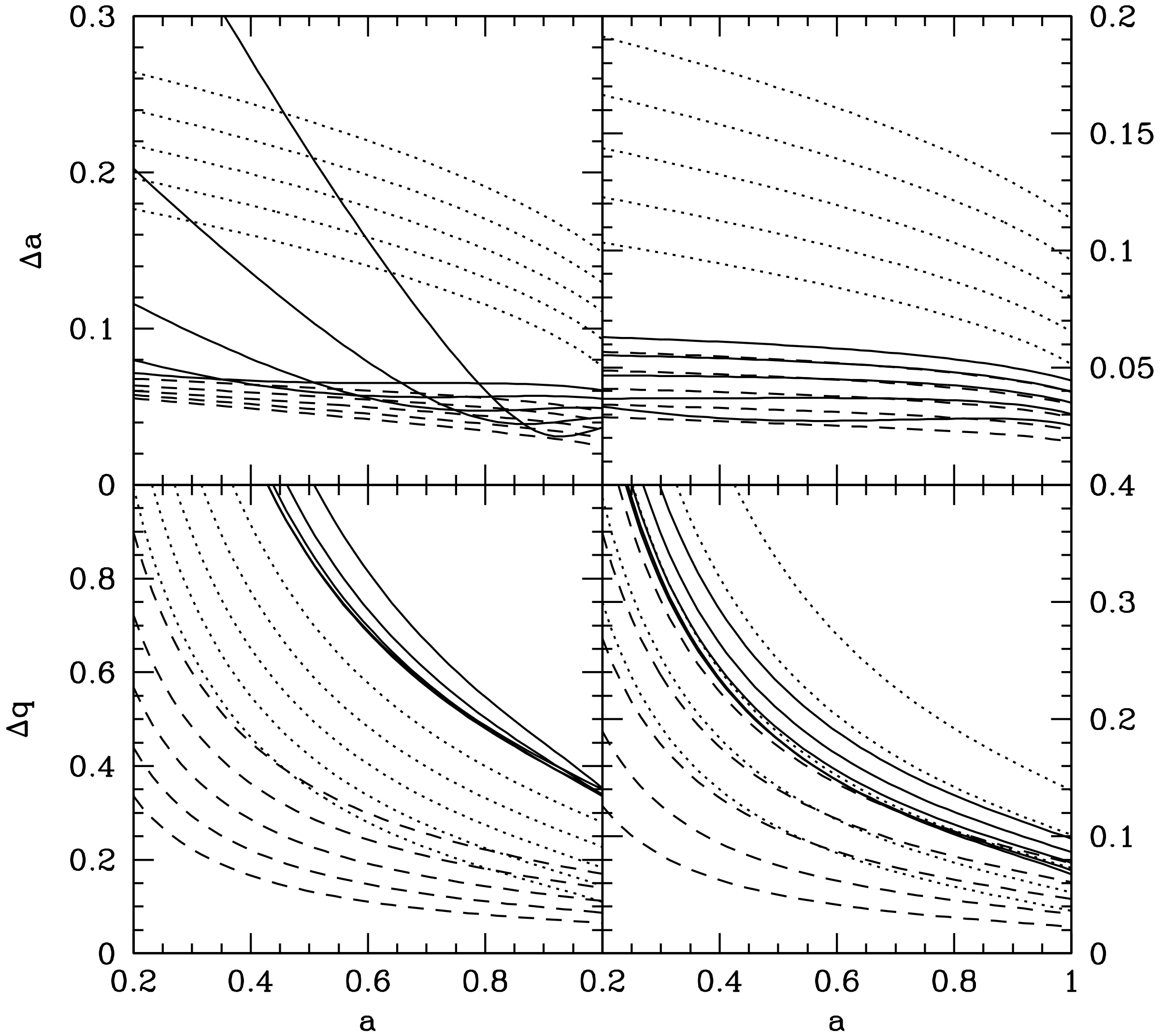

Fig. 4. The errors Δa and Δq, as defined in the text, plotted as a function of the input value of a. The lens model Family A is considered here. The left panels are for $\theta_{\rm in} = 0\farcm6$, $\theta_{\rm out} = 2\farcm0$, and the right panels for $\theta_{\rm in} = 0\farcm6$, $\theta_{\rm out} = 6\farcm0$. Each panel shows three sets of five curves; the solid curves correspond to the shear method, i.e., are based on $\langle \ell_\gamma \rangle$, the dotted curves correspond to the magnification method, and the dashed curves are the errors from using the combination of shear and magnification information. Curves within each set correspond to different input values of q, with $q = 0.6, 0.8, 1.0, 1.2, 1.4$. For $\langle \ell_\mu \rangle$ and $\langle \ell_{\rm tot} \rangle$, the curves vary monotonically with q, the lowest corresponding to $q = 0.6$, the highest to $q = 1.4$. The dependence of the errors for $\langle \ell_\gamma \rangle$ on q is more complicated and not necessarily monotonic; for small a, the uppermost curves in the left panels correspond to $q = 0.6$, for the upper right panel, the behaviour is monotonic, with Δa increasing with q. For the lower right panel, the two largest values of Δq correspond to the two largest values of q; the curves for the other values of q are almost coincident

panel), whereas for the error in the amplitude a, the relative merits of both methods depend on a and q. For the larger outer aperture radius, the amplitude of the mass distribution is much better determined from the shear method, and the relative accuracy of the slope determination depends mainly on q: for steep profiles, $q \gtrsim 1$, the shear method yields more accurate results, whereas the magnification method is superior for flatter profiles. Except for the shear method, the accuracy increases with decreasing q and increasing a, or in other words, with increasing lens strength. For the smaller outer radius of the aperture, the flattest profiles yield the largest errors for the shear method, which is counter-intuitive and must be related to the accidental cancellations of effects in the radial profile of the reduced shear, seen in Fig. 1.

6.2 Numerical simulations: comparison of the likelihood analysis with χ^2 statistics

For each method, we now quantify the agreement between the distribution of $2\Delta\langle\ell\rangle = \langle\ell\rangle(\pi) - \langle\ell\rangle(\pi_{\rm t})$ and that expected if $2\Delta\langle\ell\rangle$ followed a perfect χ^2_2 distribution. Ten thousand catalogs of lensed galaxies were generated using a Family A lens model, with true parameters $\pi_{\rm t}$ describing the cluster, and the log-likelihood functions were minimized to obtain the best fitting parameters π for each realization. For each π we then calculate $2\Delta\langle\ell\rangle$ and derive the cumulative probability distribution $P(>2\Delta\langle\ell\rangle)$, which can be compared with the perfect distribution $P(>\chi^2_2)$. In Fig. 5 the ratio of $P(>2\Delta\langle\ell\rangle) : P(>\chi^2_2)$ is plotted against $2\Delta\langle\ell\rangle$ for each of the magnification, shear and combined methods. The deviation from a χ^2_2 distribution is very small for the magnification method, and slightly larger for the shear and combined methods, but still very acceptable: the deviation from χ^2_2 statistics is such that the recovered values of π for about 94.4% of the realizations lie within the 95.4%-confidence interval.

The scatter of the recovered values of π in the a–q plane is also consistent with the ensemble-averaged log-likelihood contours, as we will demonstrate below. We can also ask how well each of the methods can distinguish *between* the Families of lens models. To these ends, we generated 500 catalogs of lensed galaxies using the prescription for numerical simulations given in Sect. 5, with a Family A lens model to describe the cluster. The log-likelihood functions ℓ_μ, ℓ_γ and $\ell_{\rm tot}$ were independently minimized under a Family A lens model and a Family B lens model, to obtain the best fitting parameters and their corresponding log-likelihood values; we refer to these as $\ell_\mu({\rm A})$, $\ell_\gamma({\rm A})$ and $\ell_{\rm tot}({\rm A})$ where the minimization is performed under a Family A lens model, and $\ell_\mu({\rm B})$, $\ell_\gamma({\rm B})$ and $\ell_{\rm tot}({\rm B})$ where a Family B lens model is used during the minimization. By comparing the log-likelihood values corresponding to recovery using each of the two Families, we determined how frequently the minimum corresponds to recovery by Family A.

In the upper, middle and lower panels of Fig. 6 we show our results for the magnification, shear and combined methods respectively. The left-hand panels show parameter recovery under a Family A lens model, whereas a Family B lens model was used to recover the best fitting parameters indicated on the right-hand panels. Contours of constant average likelihood are marked on the left-hand panel; note that the distributions of Monte Carlo points are consistent with these.

In 47% of the realizations $\ell_\mu({\rm A}) < \ell_\mu({\rm B})$ indicating that the magnification method cannot discriminate between Family A and Family B. The middle panel of Fig. 6 shows that the shear method performs slightly better: in 64% of the realizations $\ell_\gamma({\rm A}) < \ell_\gamma({\rm B})$. The results for recovery using the combined method are shown in the bottom panel; $\ell_{\rm tot}({\rm A}) < \ell_{\rm tot}({\rm B})$ in 63% of the realizations.

We repeated the same procedure, generating catalogs with a Family B lens model, and performing the minimization of ℓ_μ, ℓ_γ and $\ell_{\rm tot}$ under a Family B lens model and under a Family A lens model; our results are shown in Fig. 7. Contours of constant

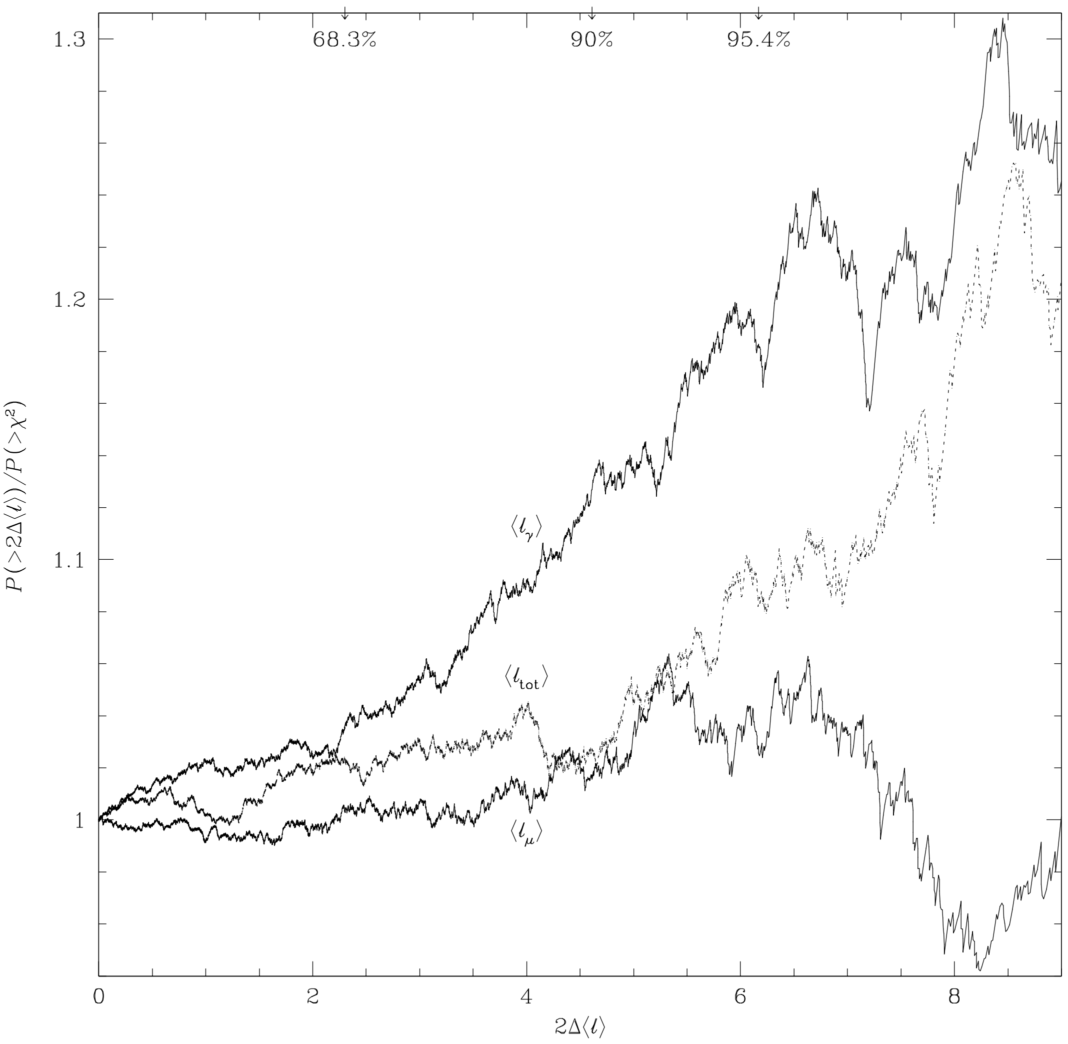

Fig. 5. For each of the magnification (lower curve), shear (upper curve) and combined methods (dashed middle curve), we show the ratio of $P(> 2\varDelta\langle\ell\rangle)$ to that expected if $2\varDelta\langle\ell\rangle$ followed a χ^2 distribution versus $2\varDelta\langle\ell\rangle$. On the top edge of the plot the 68.3%-, 90%- and 95.4%-confidence intervals are marked

average likelihood are marked on the left-hand panels and the Monte Carlo points are consistent with these.

Again, the shear method (middle panel) fares a little better than the magnification method (top panel) in discriminating between the Families: in 57% of the realizations, ℓ_μ(B)$<$ ℓ_μ(A) and in 73% of the realizations ℓ_γ(B)$<$ ℓ_γ(A). The combined method (bottom panel) gives $\ell_{\rm tot}$(B)$<$ $\ell_{\rm tot}$(A) in 78% of the realizations.

6.3 Uncertainties in the unlensed number counts

Up to now we have assumed that the number density of background galaxies at the flux threshold of the observations is known. In fact, the shape of the likelihood contours in Figs. 2 and 3 are such that they closely follow the curve of constant expected total galaxy number in the annulus, shown as the middle of the heavy dashed curves in those figures which are curves of constant $\langle N_\mu\rangle = \langle N_\mu\rangle_{\rm t}$. This implies that to first order, the

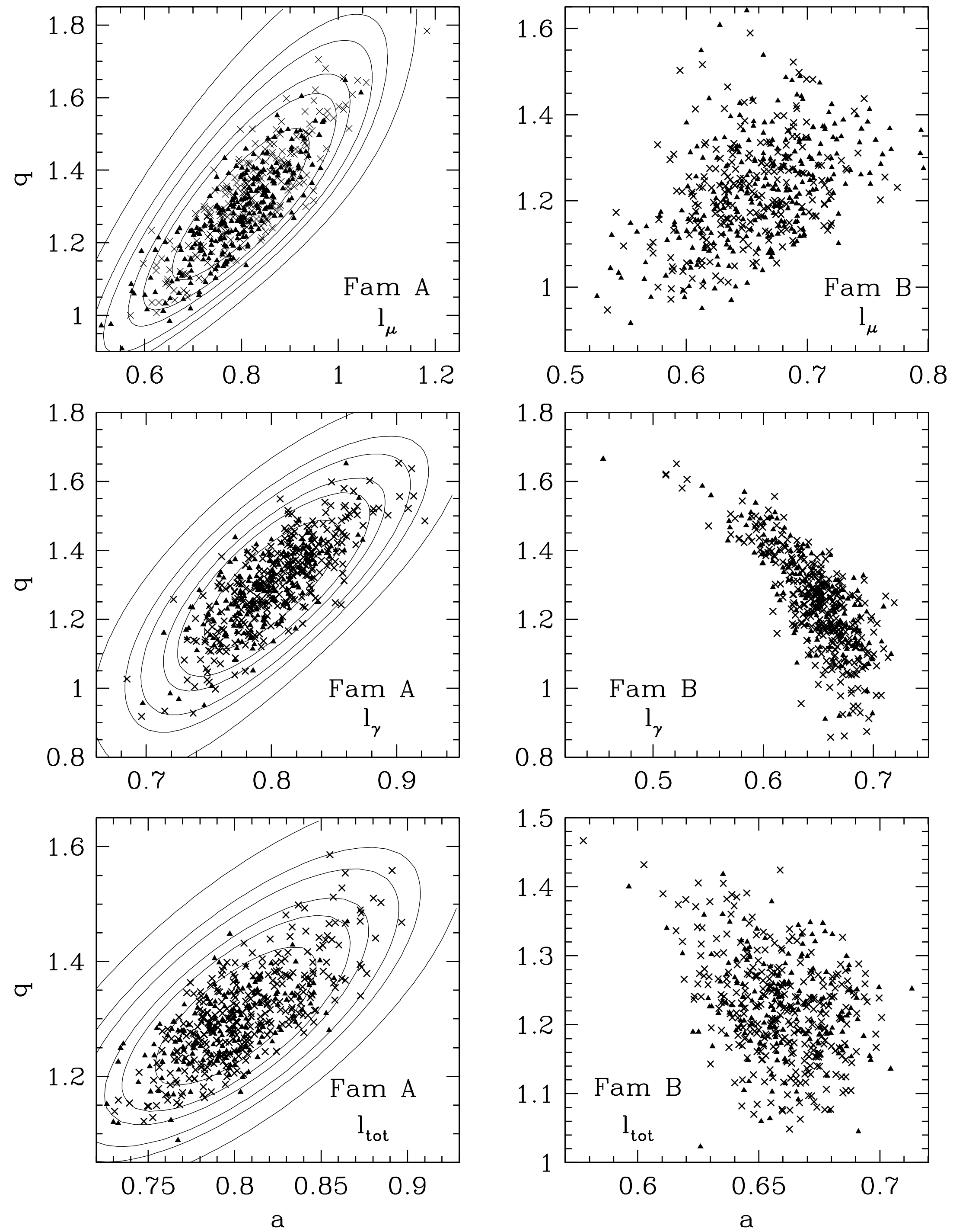

Fig. 6. A Family A input model ($a = 0.8$, $q = 1.3$) was used to generate lensed galaxy catalogs. For each of the magnification (top panel), shear (middle panel) and combined methods (bottom panel) the left-hand (right-hand) panel shows parameter recovery under a Family A (Family B) model and the crosses (triangles) indicate where the log-likelihood function is lowest when the minimization is performed using a Family A (Family B) lens model. On the left-hand panels we mark contours of constant $\Delta \langle \ell \rangle$ with levels the same as in Fig. 2

magnification information is provided by this total number. As indicated by the other two heavy solid curves in these figures, a change of the number by $\pm 5\%$ changes the curve of constant $\langle N_\mu \rangle$ quite drastically. Given that a high precision of the number density of very faint galaxies may be difficult to achieve, we shall now consider two different assumptions: first, we shall assume that the unlensed number density is not determined from any calibration frame, but is obtained from the present data set itself. After that, we shall consider the (more realistic) assumption that the unlensed number counts is known, but with some uncertainty. To this end, we reinsert the density-dependent term, which we discarded in (11), into the log-likelihood function, which then reads

$$\ell_\mu = n_\mu \, I + (1-\beta) \sum_{i=1}^{N_\mu} \ln \mu(\boldsymbol{\theta}_i) - N_\mu \, \ln n_\mu \;, \tag{40}$$

where we have defined

$$I := \int \mathrm{d}^2\theta \; [\mu(\boldsymbol{\theta})]^{\beta-1} \;. \tag{41}$$

This expression can now be minimized with respect to the unknown density n_μ, which yields, not unexpectedly,

$$n_\mu = \frac{N_\mu}{I} \;.$$

Inserting this value into (40) yields the new log-likelihood function, up to a model-independent constant,

$$\hat{\ell}_\mu = (1-\beta) \sum_{i=1}^{N_\mu} \ln \mu(\boldsymbol{\theta}_i) + N_\mu \ln I \;. \tag{42}$$

Following the techniques applied in Sect. 4, the ensemble average of $\hat{\ell}_\mu$ can be calculated,

$$\begin{aligned} \left\langle \hat{\ell}_\mu \right\rangle = {} & \bar{n}_\mu \left(\int \mathrm{d}^2\theta \; [\mu_\mathrm{t}(\boldsymbol{\theta})]^{\beta-1} \right) \ln I \\ & + \bar{n}_\mu (1-\beta) \int \mathrm{d}^2\theta \; [\mu_\mathrm{t}(\boldsymbol{\theta})]^{\beta-1} \, \ln \mu(\boldsymbol{\theta}) \;, \end{aligned} \tag{43}$$

where now $\bar{n}_\mu$ is the true number density of galaxies. By differentiation with respect to π_i, one can easily show that $\left\langle \hat{\ell}_\mu \right\rangle$ has a minimum when $\pi = \pi_\mathrm{t}$.

In Fig. 8 we have plotted contours of constant likelihood of $\left\langle \hat{\ell}_\mu \right\rangle$, for two values of the outer radius of the annulus, together with the corresponding contours of $\langle \ell_\mu \rangle$, $\langle \ell_\gamma \rangle$, and $\left\langle \hat{\ell}_\mathrm{tot} \right\rangle := \left\langle \ell_\gamma + \hat{\ell}_\mu \right\rangle$. From the figure it becomes immediately clear that dropping the assumption about the knowledge of the unlensed number density leads to a drastic loss of information from the magnification method. The thin dashed contours in Fig. 8 which correspond to $\left\langle \hat{\ell}_\mu \right\rangle$ are much wider than the solid contours corresponding to $\langle \ell_\mu \rangle$. In addition, whereas the contours of $\langle \ell_\mu \rangle$ were seen to be substantially misaligned relative to those of $\langle \ell_\gamma \rangle$, so that the combination of shear and magnification information substantially decreases the error region in parameter space relative to any single one of these methods, the contours of $\left\langle \hat{\ell}_\mu \right\rangle$ are basically parallel to those of $\langle \ell_\gamma \rangle$ (shown

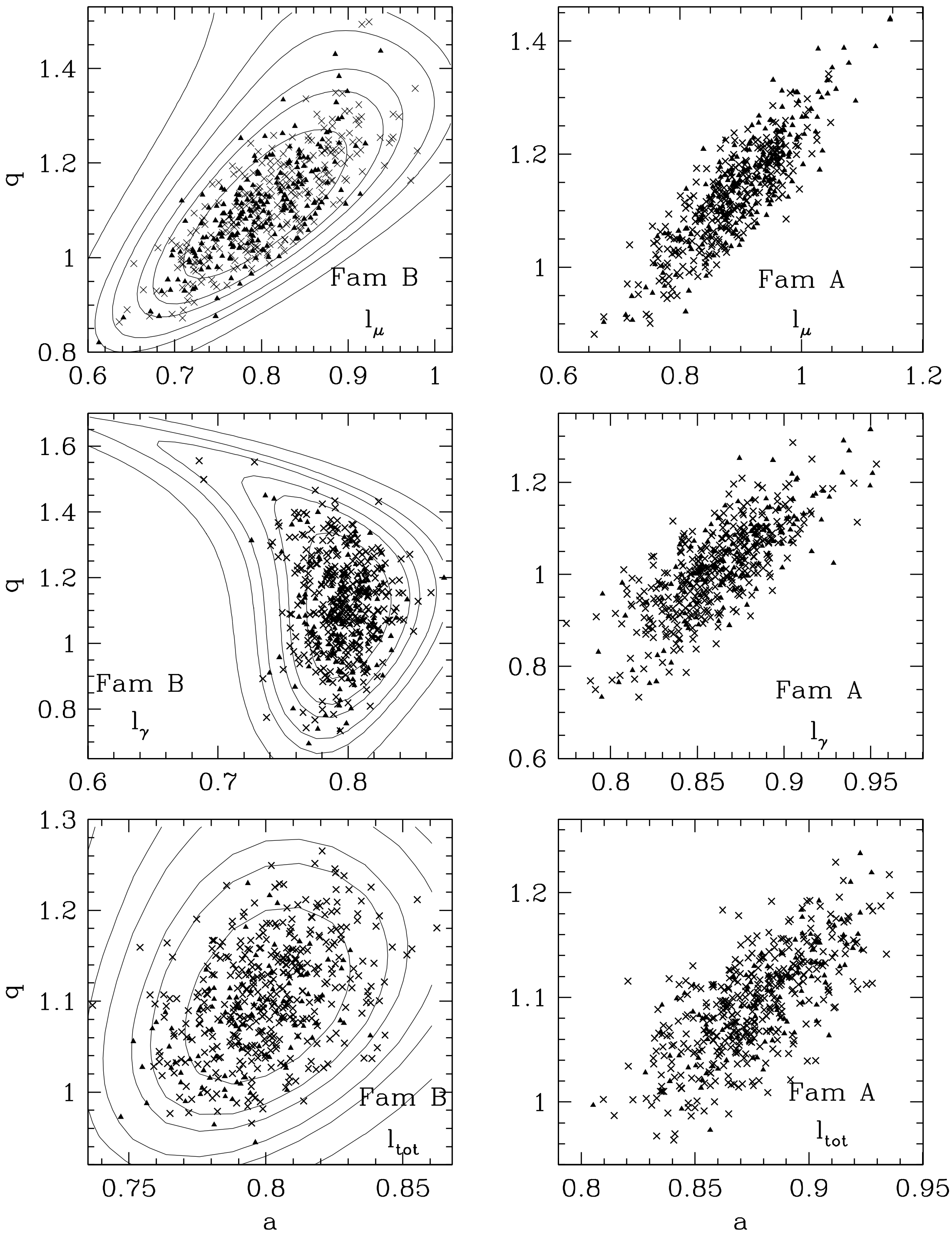

Fig. 7. A Family B input model ($a = 0.8$, $q = 1.1$) was used to generate lensed galaxy catalogs. For each of the magnification (top panel), shear (middle panel) and combined methods (bottom panel) the left-hand (right-hand) panel shows parameter recovery under a Family B (Family A) model and the crosses (triangles) indicate where the log-likelihood function is lowest when the minimization is performed using a Family B (Family A) model. On the left-hand panels we mark contours of constant $\Delta \langle \ell \rangle$ with levels the same as in Fig. 2

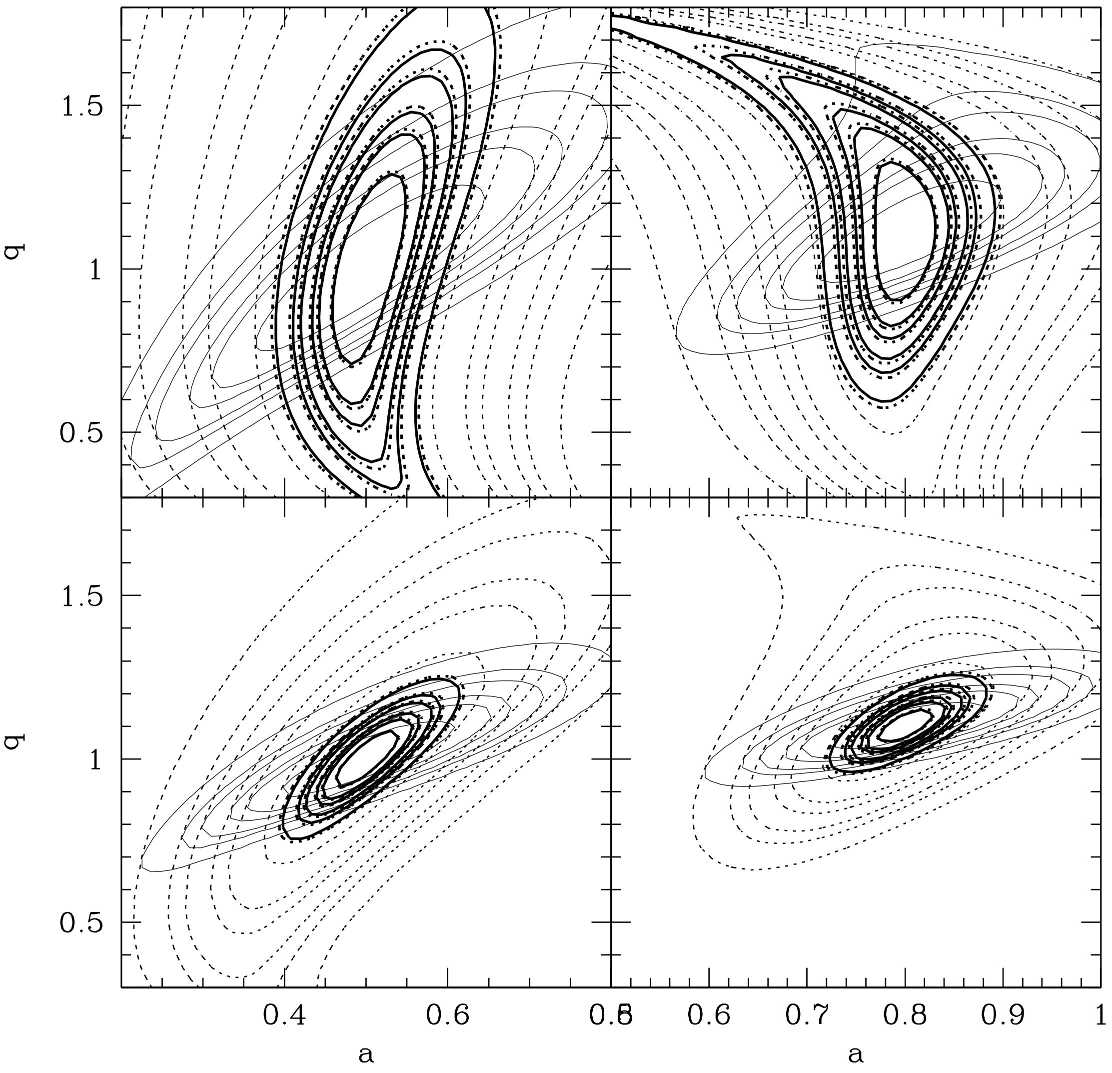

Fig. 8. Similarly to Figs. 2 and 3, contours of constant $\Delta\langle\ell\rangle$ are plotted; the contour levels are the same as described in Fig. 2. The inner radius of the annulus is $\theta_{\rm in}=0\farcm6$, the outer radius is $\theta_{\rm out}=4'$ (upper panels) and $15'$ (lower panels); left (right) panels correspond to models of Family A (Family B). The thin dashed contour correspond to $\left\langle\hat\ell_\mu\right\rangle$, the thin solid contours to $\langle\ell_\mu\rangle$, the heavy dashed contours to $\langle\ell_\gamma\rangle$, and the heavy solid contours to $\left\langle\hat\ell_{\rm tot}\right\rangle:=\left\langle\ell_\gamma+\hat\ell_\mu\right\rangle$. The input model is $a=0.5$, $q=1.0$ for Family A, and $a=0.8$, $q=1.1$ for Family B

as heavy dashed curves in Fig. 8). Therefore, the combination of both methods, in the absence of knowledge of the unlensed number density, yields only slightly more accurate parameter estimates than the shear method alone.

To understand this point, one needs to realize why the magnification information turned out to be so powerful when n_μ is known. As we have seen in Fig. 1, the reduced shear of models with different q is very similar, owing to the mass-sheet degeneracy discussed in the introduction. Knowing $\bar n_\mu$, the mass-sheet degeneracy is broken, which leads to the sensitivity of the magnification to the slope of the density profile. However, if $\bar n_\mu$ is unknown, the mass-sheet degeneracy remains, and the discussion of the relative merits of shear and magnification methods in Sect. 2 prevails.

We now make the more realistic assumption that the unlensed number density is known approximately. Suppose one knows the value of $\bar{n}_\mu$ with a fractional accuracy of η; one can then supplement the likelihood fucntion with a prior, taken to be a Gaussian in n_μ, with mean $\bar{n}_\mu$, and dispersion $\sigma_n = \eta \bar{n}_\mu$. Hence, the corresponding log-likelihood function becomes

$$\ell_\mu = n_\mu \, I + (1-\beta) \sum_{i=1}^{N_\mu} \ln \mu(\boldsymbol{\theta}_i) - N_\mu \, \ln n_\mu + \left(\frac{(n_\mu - \bar{n}_\mu)^2}{2(\eta \bar{n}_\mu)^2} \right) \; . \tag{44}$$

As before, this can be minimized with respct to n_μ, yielding

$$\frac{n_\mu}{\bar{n}_\mu} = \frac{1}{2} \left(1 - \eta^2 \bar{n}_\mu I \right) + \sqrt{\frac{1}{4} \left(1 - \eta^2 \bar{n}_\mu I \right)^2 + \eta^2 N_\mu} \; . \tag{45}$$

Inserting this value for n_μ into (44) yields the new log-likelihood function, which we shall denote as $\ell_\mu^{(\eta)}$. Taking the ensemble average in the present case is slightly more difficult, owing to the occurrence of N_μ in the square-root. However, if we neglect the Poisson-fluctuation in the ensemble averaging in (17), which is a very good approximation when $\langle N_\mu \rangle_{\rm t}$ is large, the ensemble average can be obtained as

$$\left\langle \ell_\mu^{(\eta)} \right\rangle = \langle n_\mu \rangle \, I + \bar{n}_\mu (1-\beta) \int {\rm d}^2\theta \; \mu_{\rm t}^{\beta-1} \, \ln \mu - \langle N_\mu \rangle_{\rm t} \, \ln \langle n_\mu \rangle + \frac{(\langle n_\mu \rangle - \bar{n}_\mu)^2}{2(\eta \bar{n}_\mu)^2} \; , \tag{46}$$

where

$$\frac{\langle n_\mu \rangle}{\bar{n}_\mu} = \frac{1}{2} \left(1 - \eta^2 \bar{n}_\mu I \right) + \sqrt{\frac{1}{4} \left(1 - \eta^2 \bar{n}_\mu I \right)^2 + \eta^2 \langle N_\mu \rangle_{\rm t}} \; . \tag{47}$$

Although this expression looks fairly complicated, it is easy to show that $\left\langle \ell_\mu^{(\eta)} \right\rangle$ has a minimum when π takes the value of the true model. For $\eta \to 0$, this log-likelihood reduces, up to additive constants, to the one for which the unlensed number density was assumed to be known, i.e., eq.(11), whereas for large η, it approaches (43).

In Fig. 9 we have plotted the 90%-confidence contours as obtained from $\left\langle \ell_\mu^{(\eta)} \right\rangle$, using models of Family A, for various outer radii of the data annulus. Four values of the fractional uncertainty η in the number density have been chosen, ranging from $\eta = 0$ – for which the contours agree with the corresponding ones in Fig. 2 – to $\eta = 0.06$. In addition, the 90%-confidence contour for the shear and those for the combined method are plotted. As expected, by increasing η, the confidence region increases, and the relative increase is larger, the larger $\theta_{\rm out}$. The reason for this is that for the small annulus, the relative change of the number density by magnification is much larger than the fractional uncertainty, whereas this is no longer true at larger radii. In fact, as can be seen from the lower right panel in Fig. 9, the confidence contours seem to approach an asymptotic form quickly for large annuli, implying that the prior information becomes irrelevant unless η is very small, since the unlensed number density is then more accurately determined from the data set itself. Thus, the outer-most contour in the lower-right panel in Fig. 9 is very similar to the corresponding one in Fig. 8, where no prior information about the number density was assumed. Thus, the larger the region from which data are used to determine the mass profile parameters, the more accurate the unlensed number density needs to

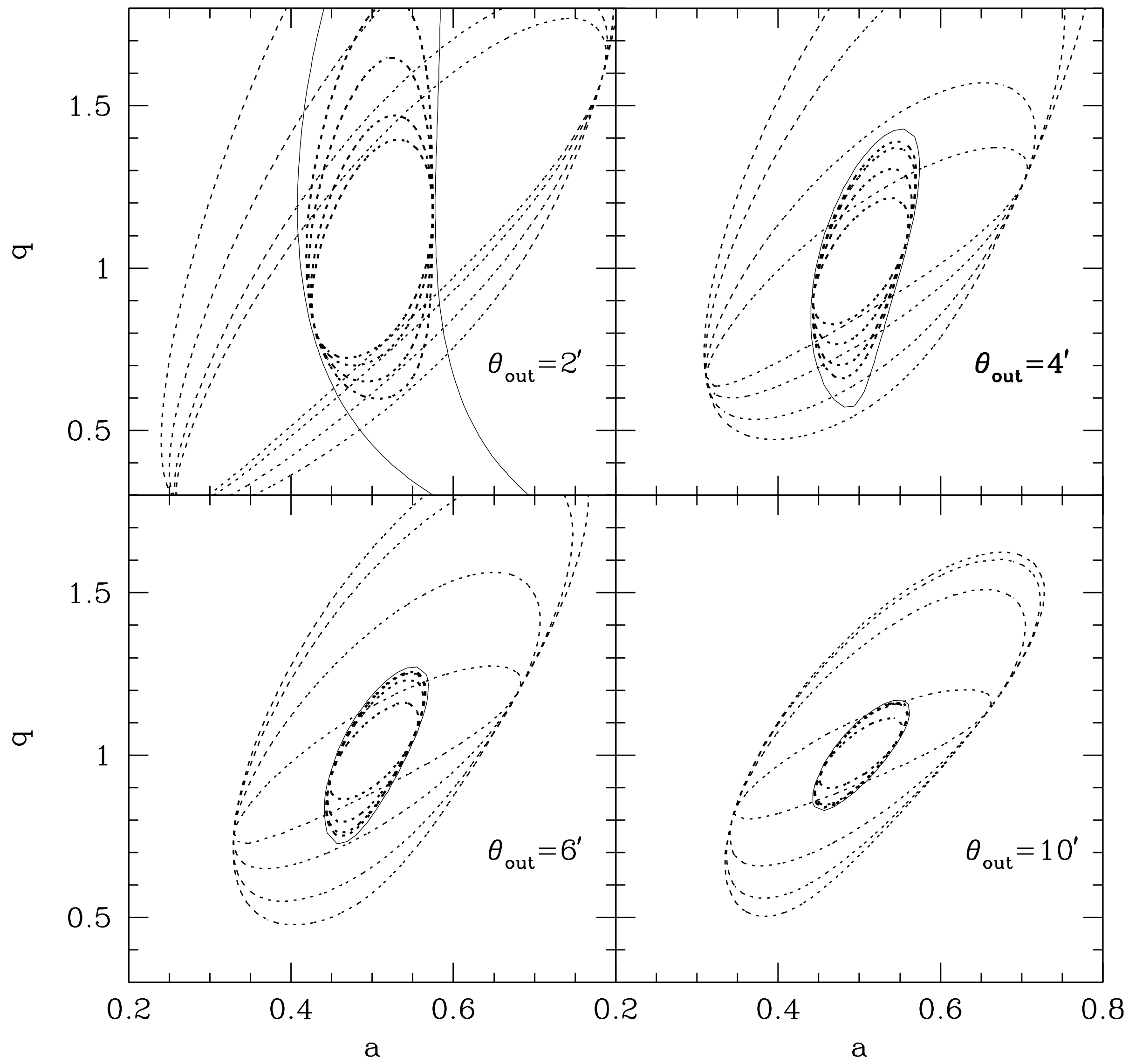

Fig. 9. For the case that the unlensed number counts are known with a fractional accuracy η, the light dashed contours plot the 90%-confidence contours of parameter estimates as obtained from the ensemble-averaged log-likelihood function (46). The model $a = 1/2$, $q = 1$ of Family A has been taken as input. The inner radius of the annulus in which data are assumed to be is $\theta_{\rm in} = 0\farcm6$, and the outer radius is different in each panel. The solid contour is the 90%-confidence contour as obtained from the shear, and the heavy dashed curves from combining shear and magnification constraints. The sets of four dashed curves are for $\eta = 0$, 0.02, 0.04, and 0.06.

be known, for a given accuracy relative to that obtainable from the shear method. In particular, we conclude that the magnification method considered here seems to be most powerful for data sets which do not extend far from the cluster center, whereas for wide-field imaging data, the shear method is superior.

Besides observational difficulties to determine the unlensed number density with a given selection function from calibration data, the anglar correlation function of galaxies provides a fundamental limit down to which the unlensed number density can be determined within any given region. To this end, we have calculated the fluctuations of the number of galaxies in a circular aperture of radius θ, using a two-point correlation function of the form $\omega(\vartheta) = A\vartheta^{-0.8}$. As a result, we find that $\delta N/N \propto \theta^{-0.4}$, a slowly decreasing function of the radius. Unfortunately, at the faint magnitude limits where the magnification method is applied, no very accurate determination of the angular two-point correlation function is available; the best measurements has been carried out in the Hubble Deep Field (e.g., Villumsen et al. 1997, and references therein). Whereas at $B \sim 28$ a clear signal in the correlation function is seen on scales below ~ 1 arcsec, the signal is too weak, and the solid angle of the HDF too small to determine whether the power-law behaviour of ω extends beyond arcsecond scale. If one extrapolates the measurements of Villumsen et al. (1997) to larger angular scales, then $\delta N/N \sim 0.1$ on an angular scale of $\sim 4'$. In that case, the results in Fig. 9 suggest that the value of the magnification method is rather limited. More likely, though, the preceeding value is a vast overestimate of the number count fluctuations due to angular clustering at these magnitudes.

7 Discussion and conclusions

Using maximum likelihood techniques, we have investigated the accuracy with which weak gravitational lensing can recover the radial density profile of clusters of galaxies. To keep the treatment as simple as possible, we have assumed that the Einstein radius of the cluster is known from strong lensing features, and that only the weak lensing information from outside the Einstein radius is used; we also confined the consideration to axially-symmetric mass distributions only. We have compared the shear method with the magnification method in form of the number depletion. Although from a simple analytical consideration the latter method is expected to yield less accurate results, the mass sheet degeneracy causes the reduced shear to be nearly degenerate along a one-dimensional subset of the two-parameter mass models considered here. In some cases it then turns out that the magnification method can constrain the parameters of the mass model more accurately than the shear method. However, this is true only as long as the unlensed number density of galaxy images is assumed to be known quite accurately. In the absence of this knowledge, the magnification method is inferior to the shear method. Since the number density of galaxies obtained from 'blank field' imaging depends not only on the flux threshold (which requires accurate photometry), but also on the seeing and sky brightness, this accurate number density calibration may be difficult to achieve. The advantage of the shear method is that no external calibration is required (the external calibration is replaced by the assumption of random intrinsic orientations). In addition, the angular correlation of faint galaxies may present a fundamental obstacle for an accurate determination of the unlensed counts in a given area, but too little is known about the angular correlation function at flux limits at which the magnification method is applied to make quantitative estimates of this effect at present.

Without accurate external calibration, our results show that large areas around the cluster need to be mapped, since the external calibration is effectively replaced by calibration from the outer region of the data field. In order for this calibration to be accurate the image quality needs to be homogeneous over the wide field. Of course, since the measurement of shear requires superb observing conditions, in particular with regards to the seeing, the number count magnification method may be applied to photometric data set which cannot be used for the shear method.

Although we have not investigated the magnification effect based on the change of image size at constant surface brightness (Bartelmann & Narayan 1995), the previous remarks on external calibration apply there as well.

External calibration may be possible using HST observations where the observing conditions are much more stable than from the ground. Hence, the claim by Broadhurst (1999) that the mass profile of the cluster Cl0024+17 can be better constrained with the magnification method than with the shear method is not contradicted by our results here; in fact, if the external calibration of the number density is possible for these HST images, the results shown in Figs. 2 and 3 indicates a higher accuracy of the shape parameter than possible with the shear method.

We would like to thank Matthias Bartelmann and Tom Broadhurst for several fruitful discussions, and also the former for carefully reading the manuscript. This work was supported by the TMR Network "Gravitational Lensing: New Constraints on Cosmology and the Distribution of Dark Matter" of the EC under contract No. ERBFMRX-CT97-0172 and the "Sonderforschungsbereich 375-95 für Astro–Teilchenphysik" der Deutschen Forschungsgemeinschaft.

Appendix

We derive here the dispersion σ of the image ellipticity distribution. As in (2) let $p_\epsilon(\epsilon^{\rm s})$ be the probability density of the intrinsic source ellipticities. Without loss of generality, we choose the reduced shear g to be real, and write $\epsilon^{\rm s} = x\mathrm{e}^{\mathrm{i}\varphi}$. Then, the transformation (1) reads

$$\begin{aligned} \epsilon_1 &= g + (1-g^2)\, x\, \frac{gx + \cos\varphi}{1 + 2gx\cos\varphi + g^2x^2} \;, \\ \epsilon_2 &= (1-g^2)\, x\, \frac{\sin\varphi}{1 + 2gx\cos\varphi + g^2x^2} \;. \end{aligned} \tag{A1}$$

Then, the dispersion of the observed ellipticities in the two directions becomes

$$\begin{aligned} \sigma_\parallel^2 &= \int_0^1 \mathrm{d}x\; x\, p(x) \int_0^{2\pi} \mathrm{d}\varphi\;\; (\epsilon_1 - g)^2 \\ &= (1-g^2)^2 \int_0^1 \mathrm{d}x\; x^3\, p_\epsilon(x) \int_0^{2\pi} \mathrm{d}\varphi\;\; \left(\frac{gx + \cos\varphi}{1 + 2gx\cos\varphi + g^2x^2}\right)^2 \\ &= (1-g^2)^2 \pi \int_0^1 \mathrm{d}x\;\, \frac{x^3 p_\epsilon(x)}{1 - x^2 g^2} = \sigma_\perp^2 \;. \end{aligned} \tag{A2}$$

References

Bartelmann, M. & Narayan, R. 1995, ApJ 451, 60

Bartelmann, M., Narayan, R., Seitz, S., Schneider, P. 1996, ApJ 464, L115

Bonnet, H., Mellier, Y. & Fort, B. 1994, ApJ 427, L83

Broadhurst, T.J., Taylor, A.N. & Peacock, J.A. 1995, ApJ 438, 49
Broadhurst, T.J. 1999, talk given at the VLT Opening Symposium, March 1–4, 1999, Antofagasta, Chile
Canizares, C.R. 1992, ApJ 263, 508
Clowe, D., Luppino, G.A., Kaiser, N., Henry, J.P. & Gioia, I. 1998, ApJ 497, L61
Fahlman, G., Kaiser, N., Squires, G. & Woods, D. 1994, ApJ 437, 56
Falco, E.E., Gorenstein, M.V., Shapiro, I.I. 1985, ApJ 289, 1L
Fischer, P. 1999, AJ 117, 2024
Fort, B., Mellier, Y. & Dantel-Fort, M. 1997, A&A 321, 353
Geiger, B. & Schneider, P. 1998, MNRAS 295, 497
Geiger, B. & Schneider, P. 1999, MNRAS 302, 118
Hoekstra, H., Franx, M., Kuijken, K., Squires, G. 1998, ApJ 504, 636
Kaiser, N. & Squires, G. 1993, ApJ 404, 441
Kaiser, N., Squires, G., Fahlmann, G. G., Woods, D. 1994, preprint CITA-94-40
Kaiser, N. 1995, ApJ 439, L1
Kaiser, N., Squires, G., Broadhurst, T. 1995, (KSB95), ApJ 449, 460
Kaiser, N., Wilson, G., Luppino, G., Kofman, L., Gioia, I., Metzger, M., Dahle, H. astro-ph/9809268, submitted to ApJ
Kochanek, C.S. 1990, MNRAS 247, 135
Lombardi, M. & Bertin, G. 1998a, A&A 335, 1
Lombardi, M. & Bertin, G. 1998b, A&A 330, 791
Luppino, G. A., Kaiser, N. 1997, ApJ 475, 20
Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. 1992, in Numerical recipes in FORTRAN. The art of scientific computing, Cambridge: University Press, 2nd Edn.
Schneider, P. 1984, A&A 140, 119
Schneider, P., Ehlers, J. & Falco, E.E. 1992, *Gravitational lenses*, Springer: New York
Schneider, P. 1995, A&A 302, 639
Schneider, P., Seitz, C. 1995, A&A 294, 411
Schramm, T. & Kayser, R. 1995, A&A 289, 5L
Seitz, C. & Schneider, P. 1995, A&A 297, 287
Seitz, C., Kneib, J.-P., Schneider, P. & Seitz, S. 1996, A&A 314, 707
Seitz, S. & Schneider, P. 1996, A&A 305, 383
Seitz, C. & Schneider, P. 1997, A&A 318, 687
Seitz, S., Schneider, P., Bartelmann, M. 1998, A&A 337, 325
Squires, G. & Kaiser, N. 1996, ApJ 473, 65
Squires, G. et al. 1996a, ApJ 461, 572
Squires, G., Kaiser, N., Fahlman, G., Babul, A. & Woods, D. 1996b, ApJ 469, 73
Taylor, A.N., Dye, S., Broadhurst, T.J., Benitez, N. & Van Kampen, E. 1998, ApJ 501, 539
Tyson, J.A., Valdes, F. & Wenk, R.A. 1990, ApJ 349, L1
Villumsen, J.V., Freudling, W. & Da Costa, L.N. 1997, ApJ 481, 578
Webster,R.L. 1985, MNRAS 213, 871